UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

CRT-606

A GENERALIZATION OF AUTOMOBILE INSURANCE RATING MODELS: THE NEGATIVE BINOMIAL DISTRIBUTION WITH A REGRESSION COMPONENT

by

Georges Dionne
and
Charles Vanasse

This research was financed by the Social Sciences and Humanities Research Council of Canada. Comments by Stuart Klugman, Jean Lemaire, and David Scott on a previous version of the paper were very useful.

Université de Montréal
Centre de recherche sur les transports - Publication #606
December 1988

Dépôt légal - 1e trimestre 1989

ABSTRACT

The objective of this paper is to provide an extension of well-known models of tarification in automobile insurance. The analysis begins by introducing a regression component in the Poisson model in order to use all available information in the estimation of the distribution. In a second step, a random variable is introduced in the regression component of the Poisson model and a negative binomial model with a regression component is derived. We then present our main contribution by proposing a bonus-malus system which integrates a priori and a posteriori information on an individual basis. We show how multivariate net premium tables can be derived from the model. Examples of tables are presented.

Key words : multivariate automobile insurance rating, Poisson model, negative binomial model, regression component, multivariate net premium tables.

RÉSUMÉ

L'objectif de cette recherche est de proposer une extension des modèles de tarification de l'assurance automobile. L'analyse débute par une présentation du modèle Poisson avec une composante de régression afin d'utiliser toute l'information disponible dans l'estimation de la distribution. Dans une seconde étape, une variable aléatoire est introduite dans la composante de régression du modèle Poisson, ce qui nous permet de dériver une distribution binomiale négative avec une composante de régression. Nous présentons alors notre principale contribution en introduisant un système bonus-malus qui intègre l'information a priori à celle a posteriori sur une base individuelle. Nous montrons comment des tableaux de primes d'assurance multivariés peuvent être obtenus du modèle. Des exemples de tableaux sont donnés.

Mots clés : tarification de l'assurance automobile multivariée, Poisson, binomiale négative, composante de régression, tableaux d'assurance multivariés.

Introduction

The objective of this paper is to provide an extension of well known models of tarification in automobile insurance. Two types of tarification are presented in the literature : 1) a priori models that select tariff variables, determine tariff classes and estimate premiums (see Eeghen et al. (1983) for a good survey of these models); 2) a posteriori models or bonus-malus systems that adjust individual premiums according to accident history of the insured (Ferreira (1974), Lemaire (1985, 1988)). This study focuses on the selection of tariff variables using multivariate regression models and on the construction of multivariate insurance tables.

The analysis begins by introducing a regression component in both the Poisson and the negative binomial models in order to use all available information in the estimation of accident distribution. We first show how the univariate Poisson model can be extended in order to estimate different individual risks (or expected number of accidents) as a function of a vector of individual characteristics. At this stage of the analysis, there is no random variable in the regression component of the model. As for the univariate Poisson model, the randomness of the extended model comes from the distribution of accidents.

In a second step, a random variable is introduced in the regression component of the Poisson model and a negative binomial model with a regression component is derived. We then present our main contribution by proposing a bonus-malus system which integrates a priori and a posteriori information on an individual basis. Multivariate net premium tables are derived and examples of tables are presented.

1. The Basic Model

1.a. Statistical Analysis

The Poisson distribution is often used for the description of random and independent events such as automobile accidents. Indeed, under well known assumptions (see Lemaire (1985) for example), the distribution of the number of accidents during a given period can be written as

$$pr(Y_i) = \frac{e^{-\lambda} \lambda^{Y_i}}{Y_i!} \tag{1}$$

where the random variable Y_i is the observed event for agent i in a given period and λ is the Poisson parameter which can be estimated by the maximum likelihood method or the method of moments.

Implicitly, (1) assumes that all the agents have the same claim frequency. A more general model allows parameter λ to vary among individuals. If we assume that this parameter is a random variable and follows a gamma distribution with parameters a and 1/b (Greenwood and Yule (1920), Seal (1969)), the distribution of the number of accidents during a given period becomes

$$pr(Y_i) = \int_0^\infty \frac{e^{-\lambda} \lambda^{Y_i}}{Y_i!} f(\lambda)\, d\lambda \tag{2}$$

where

$$f(\lambda) = \frac{(1/b)^a e^{-\lambda(1/b)} \lambda^{a-1}}{\Gamma(a)} \tag{2'}$$

is the gamma probability distribution function.
By substitution and rearrangement of the terms, (2) becomes

$$pr(Y_i) = \frac{\Gamma(Y_i+a)}{Y_i!\ \Gamma(a)} \frac{(1/b)^a}{(1+1/b)^{Y_i+a}} \tag{3}$$

which corresponds to a negative binomial distribution with $E(Y_i) = ab$ and

$Var(Y_i) = ab(1+b)$. Again, the parameters a and (1/b) can be estimated by the method of moments or by the maximum likelihood method. Empirical analyses usually reject the Poisson distribution model and yield a good representation of drivers' experience using the negative binomial distribution.

1.b. Optimal Bonus Malus Model

Following Ferreira (1974) and Lemaire (1985) an optimal bonus malus model will give the best estimator of an individual's expected number of accidents at time (t+1) given the available information for the first t periods $(Y_i^1,\ldots,Y_i^t)$. Let us denote this estimator as $\tilde{\lambda}_i^{t+1}(Y_i^1,\ldots,Y_i^t)$.

One can show that the value of the Bayes' estimator (i.e. a posteriori mathematical expectation of λ) of the true expected number of accidents for individual i is given by

$$\tilde{\lambda}_i^{t+1}(Y_i^1 \ldots\ldots Y_i^t) = \int_0^\infty \lambda\, f(\lambda/Y_i^1 \ldots\ldots Y_i^t)\,d\lambda \,. \qquad (4)$$

Applying the negative binomial distribution, the a posteriori distribution of λ is a gamma distribution with probability density function

$$f(\lambda/Y_i^1 \ldots\ldots Y_i^t) = \frac{(1/b + t)^{a+\bar{Y}_i}\, e^{-\lambda(1/b + t)}\, \lambda^{a+\bar{Y}_i-1}}{\Gamma(a+\bar{Y}_i)}, \qquad (5)$$

where $\bar{Y}_i = \sum_{j=1}^{t} Y_i^j$.

Therefore, the Bayes' estimator of an individual's expected number of accidents at time (t+1) is the mean of the a posteriori gamma distribution with parameters $(a+\bar{Y}_i)$ and (1/b+t):

$$\tilde{\lambda}_i^{t+1}\ (Y_i^1, \ldots\ldots\ldots, Y_i^t) = \frac{a + \bar{Y}_i}{(1/b)+t} \quad . \qquad (6)$$

Actuarial net premium tables can then be calculated by using (6).

2. The Generalized Model

Since past experience cannot, in a short length of time, generate all the statistical information that permits fair insurance tarification, many insurers use both a priori and a posteriori tarification systems. A priori classification is based on significant variables that are easy to observe namely, age, sex, type of driver's license, place of residence, etc. A posteriori information is then used to complete a priori classification. However, since both steps of the analysis are not adequately integrated into a single model, inconsistencies are produced. For example, in a priori classification of risks, linear regression models are estimated in order to identify classification variables. These models assume a normal distribution of accidents while, in a posteriori models, accidents are assumed to follow either Poisson or negative binomial distributions.

In this section of the paper, a priori information is integrated into Poisson and negative binomial models in order to calculate multivariate insurance tables.

2.a. Statistical Analysis

Let us begin with the Poisson model. As in the preceding section, the random variables Y_i are independent. In the extended model, however, λ may vary between individuals. Let us denote by λ_i the expected number of accidents corresponding to individuals of type i. This expected number is determined by k exogenous variables or characteristics $x_i=(x_{i1}, x_{i2}, \ldots\ldots, x_{ik})$ which represent different a priori classification variables. We can write

$$\lambda_i = \exp(x_i\beta) \qquad (7)$$

where β is a vector of coefficients (kxl). (7) implies the non-negativity of λ_i .

The probability specification becomes

$$Pr(Y_i) = \frac{e^{-\exp(x_i\beta)}(\exp(x_i\beta))^{Y_i}}{Y_i!} \quad . \qquad (8)$$

It is important to note that λ_i is not a random variable. The model assumes implicitly that the k exogenous variables provide enough information to obtain the appropriate values of the individual's probabilities. As for the univariate version of the model, the randomness is due to the Poisson specification. The β parameters can be estimated by the maximum likelihood method (see Hausman, Hall and Griliches (1984) for an application to the patents - R & D relationship). Since the model is assumed to contain all the necessary information required to estimate the values of the λ_i, there is no room for a posteriori tarification in the extended Poisson model. Finally, it is easy to verify that (1) is a particular case of (8).

The main advantage of the modified Poisson model (i.e. containing a regression component) over the linear regression model lies in the fact that the dependent variable is a count variable restricted to have only non-negative values. Consequently, the estimated expected number of accidents is clearly non-negative which is not always the case in linear regression models. Moreover, by definition, ordinary least squares imply that the marginal effect of each explanatory variable is constant and independent of other variables. Finally, individual accident data may cause heteroskedasticity and usual methods of correction for linear models could affect the results.

However, when the vector of explanatory variables does not contain all the significant information, a random variable has to be introduced into the regression component. Following Gourieroux, Monfort and Trognon (1984), (7) may be written as

$$\lambda_i = \exp(x_i\beta+\epsilon_i) \tag{9}$$

yielding a random λ_i. Equivalently, (9) can be rewritten as

$$\lambda_i = \exp(x_i\beta)u_i \tag{10}$$

where $u_i \equiv \exp(\epsilon_i)$.

As for the univariate negative binomial model presented above, if we assume that u_i follows a gamma distribution with $E(u_i)=1$ and $Var(u_i)=1/a$, the probability specification becomes

$$pr(Y_i) = \frac{\Gamma(Y_i+a)}{Y_i!\ \Gamma(a)} \left[\frac{\exp(x_i\beta)}{a}\right]^{Y_i} \left[1 + \frac{\exp(x_i\beta)}{a}\right]^{-(Y_i+a)} \tag{11}$$

which is also a negative binomial distribution with parameters a and $\exp(x_i\beta)$. We will show later that the above parameterization does not affect the results if there is a constant term in the regression component.

Then $E(Y_i) = \exp(x_i\beta)$ and $Var(Y_i) = \exp(x_i\beta)\left[1 + \frac{\exp(x_i\beta)}{a}\right]$.

We observe that $Var(Y_i)$ is a nonlinear increasing function of $E(Y_i)$. When the regression component is a constant c, $E(Y_i) = \exp(c)$ and

$Var(Y_i) = \exp(c)\left[1 + \frac{\exp(c)}{a}\right]$ which correspond, respectively, to the mean and variance of the univariate negative binomial distribution with parameters (a,1/b) where $b = \exp(c)/a$.

Dionne and Vanasse (1988) estimated the parameters of both the Poisson and negative binomial distributions with a regression component. A priori information was measured by variables such as age, sex, number of years with a driver's license, place of residence, driving restrictions, class of driver's license and number of days the driver's license was valid. The Poisson distribution with a regression component was rejected and the negative binomial distribution with a regression

component yielded better results than the univariate negative binomial distribution (see Section 3 for more details).

An extension of the Bayesian analysis was then undertaken in order to integrate a priori and a posteriori tarification models.

2.b. A Generalization of the Optimal Bonus Malus Model

Consider again an insured driver i with an experience over t periods; let Y_i^j represent the number of accidents in period j and x_i^j, the vector of the k characteristics observed at period j, that is $x_i^j = (x_{i1}^j, \ldots\ldots, x_{ik}^j)$. Let us further suppose that the true expected number of accidents of individual i at period j, $\lambda_i^j(u_i, x_i^j)$, is a function of both individual characteristics x_i^j and a random variable u_i. The insurer needs to calculate the best estimator of the true expected number of accidents at period t+1. Let $\bar{\lambda}_i^{t+1}(Y_i^1, \ldots\ldots, Y_i^t; x_i^1, \ldots\ldots, x_i^{t+1})$ designate this estimator which is a function of past experience over the t periods and of known characteristics over the t+1 periods.

If we assume that u_i are independent and identically distributed over time and that the insurer minimizes a quadratic loss function, one can show that the optimal estimator is equal to:

$$\bar{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1}) = \int_0^\infty \lambda_i^{t+1}(u_i, x_i^{t+1})\, f(\lambda_i^{t+1}/Y_i^1, \ldots, Y_i^t; x_i^1, \ldots\ldots, x_i^t)\, d\lambda_i^{t+1} \,, \qquad (12)$$

where, by the Bayes' rule,

$$f(\lambda_i^{t+1}/Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^t) = \frac{P((Y_i^1, \ldots, Y_i^t)/\lambda_i^{t+1}; x_i^1, \ldots\ldots, x_i^t) f(\lambda_i^{t+1})}{\bar{P}((Y_i^1, \ldots\ldots, Y_i^t)/x_i^1, \ldots\ldots, x_i^t)} \qquad (13)$$

and where, by definition,

$$\bar{P}((Y_i^1,\ldots,Y_i^t)/x_i^1,\ldots,x_i^t)=\int_0^\infty P((Y_i^1,\ldots,Y_i^t)/\lambda_i^{t+1};x_i^1,\ldots,x_i^t)\ f(\lambda_i^{t+1})d\lambda_i^{t+1} \tag{14}$$

Applying the negative binomial distribution to the model, the probability of the sequence $(Y_i^1,\ldots,Y_i^t)$ given the true expected number of accidents at t+1 and the individual's characteristics is a t-dimension Poisson distribution

$$P((Y_i^1,\ldots,Y_i^t)/\lambda_i^{t+1};\ x_i^1,\ldots,x_i^t) = \frac{e^{-\sum_{j=1}^{t}\lambda_i^j}\ \prod_{j=1}^{t}(\lambda_i^j)^{Y_i^j}}{\prod_{j=1}^{t}(Y_i^j!)} \tag{15}$$

where $\lambda_i^j = \exp(x_i^j\beta)u_i \equiv (\dot{\lambda}_i^j)u_i$. The unconditional distribution of $(Y_i^1,\ldots,Y_i^t)$ when u_i follows a gamma distribution with mean equal to one and variance equal to 1/a is given by

$$\bar{P}\ (Y_i^1,\ldots,Y_i^t/x_i^1,\ldots,x_i^t) = \frac{\Gamma(a+\bar{Y}_i)\ a^a\ \prod_{j=1}^{t}(\dot{\lambda}_i^j)^{Y_i^j}}{\prod_{j=1}^{t}(\lambda_i^j!)\ \Gamma(a)\ (a+\bar{\lambda}_i)^{\bar{Y}_i+a}} \tag{16}$$

where $\bar{Y}_i = \sum_{j=1}^{t} Y_i^j$ and $\bar{\lambda}_i = \sum_{j=1} \dot{\lambda}_i^j$.

Using Bayes's rule, one verifies that

$$f(\lambda_i^{t+1}/Y_i^1,\ldots,Y_i^t;x_i^1,\ldots,x_i^t) = \frac{[a+\bar{\lambda}_i]^{\bar{Y}_i+a}\ e^{-u_i(a+\bar{\lambda}_i)}\ u_i^{a-1+\bar{Y}_i}}{\Gamma(a+\bar{Y}_i)} \tag{17}$$

which corresponds to a gamma distribution with parameters $(a+\bar{Y}_i,\ a+\bar{\lambda}_i)$.

Therefore, the Bayes' optimal estimator of the true expected number of accidents for individual i is:

$$\tilde{\lambda}_i^{t+1}(Y_i^1,\ldots,Y_i^t;x_i^1,\ldots,x_i^{t+1}) = \hat{\lambda}_i^{t+1}\left[\frac{a+\bar{Y}_i}{a+\bar{\lambda}_i}\right] \qquad (18)$$

When t=0, $\tilde{\lambda}_i^1=\hat{\lambda}_i^1 \equiv \exp(x_i^1\beta)$ which implies that only a priori tarification is used in the first period. Moreover, when the regression component is limited to a constant c, one obtains:

$$\tilde{\lambda}_i^{t+1}(Y_i^1,\ldots,Y_i^t) = \exp(c)\left[\frac{a+\bar{Y}_i}{a+t\,\exp(c)}\right] = \frac{a+\bar{Y}_i}{(1/b)+t} \qquad (19)$$

which is (6). This result is not affected by the parameterization of the gamma distribution.

Two criteria define an optimal bonus-malus system which has to be fair for the policyholders and be financially balanced for the insurer. It is clear that the estimator proposed in (18) is fair since it allows the estimation of the individual's risk in function of both his characteristics and past experience. Appendix I shows that the extended model is also financially balanced.

3. Examples of multivariate tables

As mentioned above, Dionne and Vanasse (1988) have estimated the parameters of the negative binomial distribution by using a sample of about 19,000 individuals from Quebec. Many a priori variables were found significant. For example, the age and sex interaction variables were significant as well as classes of driver's licences for bus, truck, and taxi drivers. Even if the Poisson model gave similar results to those of the negative binomial model, it was shown (standard likelihood ratio test) that there was a gain in efficiency by using a model allowing for overdispersion of the data (where the variance is greater than the mean).

The usual χ^2 test generated similar conclusion. The latter results are summarized in Table 1:

Table 1

Estimates of Multivariate Poisson and Negative Binomial Distributions

Individual's number of accidents in a given period	Observed numbers of individuals during 1982-1983	Predicted numbers of individuals for 1982-1983	
		Poisson*	Negative binomial*
0	17,784	17,747.81	17,786.39
1	1,139	1,201.59	1,131.05
2	79	60.56	86.21
3	9	2.88	8.18
4	2	.15	.98
5+	0	0	0
	19,013	χ^2 = 29.91	χ^2 = 1.028
		$\chi^2_{2,.95}$ = 5.99	$\chi^2_{1,.95}$ = 3.84
		LLF = -4,661.57	LLF = -4,648.58

* The estimated parameters are published in Dionne-Vanasse (1988) and are available upon request.

The univariate model was also estimated for the purpose of comparison. The estimated parameters of the univariate model are $\hat{a}$ = .696080 and $(1/\hat{b})$ = 9.93580. Table 2 presents the results:

Table 2

Estimates of Univariate Poisson and Negative Binomial Distributions

Individual's number of accidents in a given period	Observed numbers of individuals during 1982-1983	Predicted numbers of individuals for 1982-1983	
		Poisson ($\exp \bar{c} = 0.0701$)	Negative binomial ($\hat{a}=0.6960; 1/\hat{b}=9.9359$)
0	17,784	17,726.60	17,785.28
1	1,139	1,241.86	1,132.05
2	79	43.50	88.79
3	9	1.02	7.21
4	2	0.02	.61
5+	0	0	0
	19,013	$\chi^2 = 133.06$	$\chi^2 = 2.21$
		$\chi^2_{2,.95} = 5.99$	$\chi^2_{1,.95} = 3.84$
		LLF = -4950.28	LLF = -4916.78

Using the values of the estimated parameters of the univariate negative binomial distribution presented above, Table 3 was formed by applying (20) where \$100 is the first period premium:

$$\hat{P}_i^{t+1}(Y_i^1,\ldots,Y_i^t) = 100\,\frac{1/\hat{b}\,(\hat{a} + \bar{Y}_i)}{\hat{a}\,(1/\hat{b} + t)} \qquad (20)$$

Table 3

Univariate Negative Binomial Model

$\hat{a} = .696080 \quad 1/\hat{b} = 9.93580$

t \ $\bar{Y}_t$	0	1	2	3	4
0	100.00				
1	90.86	221.38	351.91	462.43	612.96
2	83.24	202.83	322.42	442.01	561.60
3	76.81	187.15	297.50	407.84	518.19
4	71.30	173.72	276.15	378.58	481.00
5	66.52	162.09	257.66	353.23	448.80
6	62.35	151.92	241.49	331.06	420.63
7	58.67	142.95	227.23	311.52	395.80
8	55.40	134.98	214.56	294.15	373.73
9	52.47	127.85	203.23	278.61	353.99

In Table 3, we observe that only two variables may change the level of insurance premiums, i.e. time and the number of accumulated accidents. For example, an insured who had three accidents in the first period will pay a premium of \$462.43 in the next period, but if he had no accidents, he would have paid only \$90.86.

From (20) it is clear that no additional information can be obtained in order to differentiate an individual's risk. However, from (18), a more general pricing formula can be derived:

$$\hat{P}_i^{t+1}(Y_i^1 \ldots Y_i^t; x_i^1 \ldots x_i^{t+1}) = M\,\hat{\lambda}_i^{t+1}\left[\frac{\hat{a}+\bar{Y}_i}{\hat{a}+\bar{\lambda}_i}\right] \qquad (21)$$

where $\hat{\lambda}_i^{t+1} \equiv \exp(x_i^{t+1}\hat{\beta})$, $\bar{\lambda}_i \equiv \sum_{j=1}^{t}\exp(x_i^j\hat{\beta})$,

and M is such that $1/I\sum_{j=1}^{I}\hat{\lambda}_i^{t+1} M = \$100$ when the total number of insureds is I.

This general pricing formula is function of time, the number of accumulated accidents and the individual's significant characteristics in the regression component. In consequence, tables can now be constructed more generally by using (21). First, it is easy to verify that each agent does not start with a premium of $100. In Table 4, for example, a young driver begins with $280.89. Second, since the age variable is negatively significant in the estimated model, two factors, rather than one, have a negative effect on the individual's premiums (i.e. time and age). In Table 4, the premium is largely reduced when the driver reaches period seven at 25 years old (a very significant result in the empirical model).

Table 4

Negative Binomial Model With a Regression Component

Male, 18 years old in period 0, region 9, class 42

t \ $\bar{Y}_t$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	180.06	302.78	425.50	548.23	670.95
5	165.81	278.81	391.82	504.82	617.83
6	153.64	258.36	363.07	467.79	572.50
7	79.85	134.28	188.70	243.12	297.55
8	76.92	129.35	181.77	234.19	286.62
9	74.20	124.76	175.33	225.90	276.46

For the purpose of comparison, Table 4 was normalized such that the agent starts with a premium of $100. The results are presented in Table 5. The effect of using a regression component is directly observed. Again the difference between the corresponding premiums in Table 3 and Table 5 come from two sources: the individual in Table 5 has particular a priori characteristics while all individuals are implicitly assumed identical in Table 3 and age is significant when the individual reaches period seven (25 years old).

Table 5

Table 4 Divided by 2.8089

t \ $\bar{Y}_t$	0	1	2	3	4
0	100.00				
1	88.17	148.27	208.36	268.46	328.55
2	77.42	130.18	182.94	235.70	288.46
3	70.13	117.93	165.73	213.53	261.33
4	64.10	107.79	151.48	195.18	238.87
5	59.03	99.26	139.49	179.72	219.95
6	54.70	91.98	129.26	166.54	203.82
7	28.43	47.81	67.18	86.55	105.93
8	27.38	46.05	64.71	83.37	102.04
9	26.42	44.42	62.42	80.42	98.42

Moreover, when the insured modifies significant variables, new tables may be formed. In Table 4 the driver was in region # 9 (a risky region in Quebec) and had a standard driving license.

Now if the individual moves from region #9 to a less risky region (Montreal, for example) in period 4, the premiums then change (see Table 6).

Table 6

Negative Binomial Model With a Regression Component

Same individual as in Table 2, moved to Montreal in period 4

t \ $\bar{Y}_i$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	119.65	201.19	282.73	364.28	445.82
5	113.18	190.32	267.45	344.59	421.73
6	107.38	180.56	253.74	326.92	400.11
7	56.98	95.81	134.65	173.48	212.32
8	55.47	93.28	131.08	168.89	206.69
9	54.04	90.87	127.70	164.53	201.36

Having two accidents, he now pays $282.73 in period 4 instead of $425.50. Finally, if the driver decides to become a Montreal taxi driver in period 5, the following results can be seen in Table 7.

Table 7

Negative Binomial Model With a Regression Component

Same individual as in Table 4, moved to Montreal in period 4, changed for class 31 (taxi) in period 5.

t \ $\bar{Y}_t$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	119.65	201.19	282.73	364.28	445.82
5	291.65	490.42	689.19	887.96	1086.73
6	256.00	430.48	604.95	779.42	953.90
7	127.26	213.99	300.72	387.45	474.18
8	119.97	201.73	283.49	365.25	447.02
9	113.47	190.80	268.13	345.47	422.80

Again, having two accidents, he now pays \$689.19 in period 5 instead of \$267.45.

Concluding Remarks

In this paper, we have proposed an extension of well-known models of tarification in automobile insurance. We have shown how a bonus-malus system, based only on a posteriori information, can be modified in order to take into account a priori information on an individual basis. Consequently, we have integrated two well-known systems of tarification into a unified model and reduced some problems of consistencies. However, we have limited our analysis to the optimality of the model.

One line of research is the integration of accident severity into the general model. Recent contributions have analyzed different types of distribution functions to be applied to the severity of losses (Lemaire (1985) for automobile accidents, Cummins et al. (1988) for fire losses, and Hogg and Klugman (1984) for many other applications). However, to our knowledge, no study has considered the possibility of using a regression component in order to verify whether the individual's characteristics or actions are significant in accident severity. It is important to investigate that such variables are significant, in order to both build a fair rating system and increase incentives for safety.

BIBLIOGRAPHY

Cummins J.D., Dionne G., McDonald J.B. and Pritchett B.M. (1988) "Application of the GB2 Distribution in Modeling Insurance Loss Processes" Mimeo, Department of Insurance, University of Pennsylvania.

Dionne G. and Vanasse C. (1988) "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information" Mimeo, CRT and Economics Department, Université de Montréal.

Ferreira J. (1974) "The Long-Term Effects of Merit Rating Plans on Individual Motorists" Operations Research 22, 954-978.

Gourieroux C, Monfort A. and Trognon A. (1984) "Pseudo Maximum Likehood Methods: Application to Poisson Models" Econometrica, 52, 701-720.

Greenwood M. and Yule G.U. (1920) "An Inquiry into the Nature of Frequency Distribution of Multiple Happenings "Journal of Royal Statistical Society A, 83, 255-279.

Hausman J., Hall B.H. and Griliches Z. (1984) "Econometric Models for Count Data with an Application to the Patents - R & D Relationship" Econometrica, 52, 910-938.

Hogg R.V. and Klugman S.A. (1984) Loss Distributions. Wiley, New-York.

Lemaire J. (1985) Automobile Insurance: Actuarial Models. Kluwer-Nighoff, Boston, 248 pages.

Lemaire J. (1988) "A Comparative Analysis of Most European and Japanese Bonus-Malus Systems" Mimeo, Insurance Department, University of Pennsylvania. Forthcoming in the Journal of Risk and Insurance.

Seal H.L. (1969) Stochastic Theory of a Risk Business. Wiley, New-York.

Van Eeghen J., Greup E.K. and Nijssen J.A. (1983) Surveys of Actuarial Studies: Rate Making. Research Department, Nationale - Nederlanden N.V., 138 pages.

Address: Georges Dionne, CRT, Université de Montréal, C.P. 6128, Succ. A, Montreal, H3C 3J7, Canada.

Appendix

Proof that the Generalized Model is Financially Balanced

We have to show that

$$\int_0^\infty \sum_{Y_i^j}^\infty \bar{\lambda}_i^{t+1}(Y_i^1 \ldots Y_i^t; x_i^1 \ldots x_i^{t+1})\, P((Y_i^1 \ldots Y_i^t)/\lambda_i^{t+1}; x_i^1 \ldots x_i^{t+1}) f(\lambda_i^{t+1}) d\lambda_i^{t+1} \tag{A.1}$$

is equal to $\exp(x_i^{t+1}\beta) \equiv \dot{\lambda}_i^{t+1}$.

(A.1) can be rewritten as

$$\int_0^\infty \sum_{Y_i^j}^\infty \dot{\lambda}_i^{t+1} \left[\frac{\bar{a+Y_i}}{\bar{a+\lambda_i}} \right] P((Y_i^{t+1} \ldots Y_i^t)/\lambda_i^{t+1}; x_i^1 \ldots x_i^{t+1}) f(\lambda_i^{t+1}) d\lambda_i^{t+1} \tag{A.2}$$

or as

$$\dot{\lambda}_i^{t+1} \left[\frac{a}{\bar{a+\lambda_i}} \right] + \frac{\dot{\lambda}_i^{t+1}}{\bar{a+\lambda_i}} \int_0^\infty \sum_{Y_i^j}^\infty \sum_r^t Y_i^r\, P((Y_i^{t+1} \ldots Y_i^t)/\lambda_i^{t+1}; x_i^1 \ldots x_i^{t+1})\, f(\lambda_i^{t+1}) d\lambda_i^{t+1} \tag{A.3}$$

More explicitly, (A.3) can be written as

$$\dot{\lambda}_i^{t+1} \left[\frac{a}{\bar{a+\lambda_i}} \right] + \frac{\dot{\lambda}_i^{t+1}}{\bar{a+\lambda_i}} \int_0^\infty \sum_{Y_i^j}^\infty \sum_r^t Y_i^r \prod_{j=1}^t \frac{e^{\lambda_i^j} (\lambda_i^j)^{x_i^j}}{Y_i^j!} f(\lambda_i^{t+1}) d\lambda_i^{t+1}$$

and after some rearrangements we obtain

$$\dot{\lambda}_i^{t+1}\left[\frac{a}{a+\bar{\lambda}_i}\right]+\frac{\dot{\lambda}_i^{t+1}}{a+\bar{\lambda}_i}\sum_{r}^{t}\int \lambda_i^{r}\, f(\lambda_i^{t+1})d\lambda_i^{t+1} \,. \qquad \text{(A.4)}$$

(A.4) is equivalent to

$$\dot{\lambda}_i^{t+1}\left[\frac{a}{a+\bar{\lambda}_i}+\frac{\bar{\lambda}_i}{a+\bar{\lambda}_i}\right] \qquad \text{(A.5)}$$

which is the desired result.